KU LEUVEN

COSIC

Hardware: an essential partner to cryptography

IACR Distinguished Lecture, UPDATE
Prof. Dr. Ir. Ingrid Verbauwhede
KU Leuven, COSIC

erc

Eurocrypt 2022 – June 1 2022
Montreux – April 20, 2023

Slide acknowledgement: All past and present PhD students!

Outline

- Position of cryptography in the design of embedded systems
  - Root of trust & secure composition
- Cryptography relies on hardware because it needs:
  - Feasibility & Performance
  - Secure implementation: protection against side-channel, fault attacks
  - Secure key storage (PUFs)
  - Quality random number generators
  - Acceleration of new crypto: COED and FHE
- Conclusions

2

KU LEUVEN

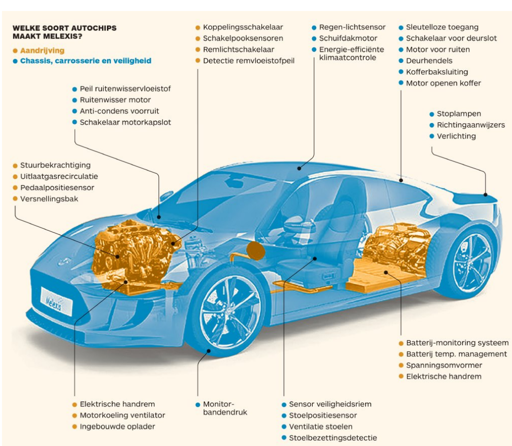
NEXT GENERATION EMBEDDED SYSTEMS

3

KU LEUVEN

Automotive

"Networked embedded systems interacting with the environment"



Today 58 Melexis chips in TESLA Model Y,
170 Melexis chips in Mercedes EQS

[De Tijd, February 2, **2022**]


ANDY GREENBERG SECURITY 07.21.15 06:00 AM

HACKERS REMOTELY KILL A JEEP ON THE HIGHWAY—WITH ME IN IT

- Networked ➔ secure, authenticated communication , low latency
- Embedded ➔ compact (no external memory), cheap, no batch processing
- Interacting with environment ➔
  - LOW latency
  - Compact
- Resistant to attacks

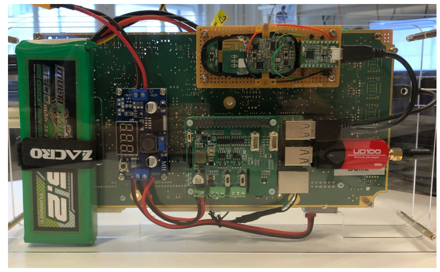
4

KU LEUVEN

How to evaluate security? Where to start?

Tesla Model X key fob (2020)
https://youtu.be/clrNuBb3myE

Tesla Model S key fob (2018)
https://youtu.be/aVlYuPzmJoY



Passive Keyless Entry and Start System:
- Wireless challenge response system
- No Mutual authentication (model S)
- Weak crypto (model S)
- Secure element, but problems with protocol (model X)
- Off the shelf radios and components

[Lennert Wouters, COSIC]

5

KU LEUVEN

TRUST AND TRUST BOUNDARIES

6

KU LEUVEN

Trust Definition

Trust (R. Anderson in "Security Engineering", after NSA):

- "Trusted system or component is one whose failure can break the security policy, while a *trustworthy* system or component is one that won't fail."

Trust (Trusted Computing Group):

- "An entity can be trusted if it always behaves in the expected manner for the intended purpose."

Loosely stated: if trusted system or component fails, then bad things can happen.

Goal of security: **minimize** what needs to be trusted

Focus on cryptography in this context?

7

KU LEUVEN

What is the root of trust?

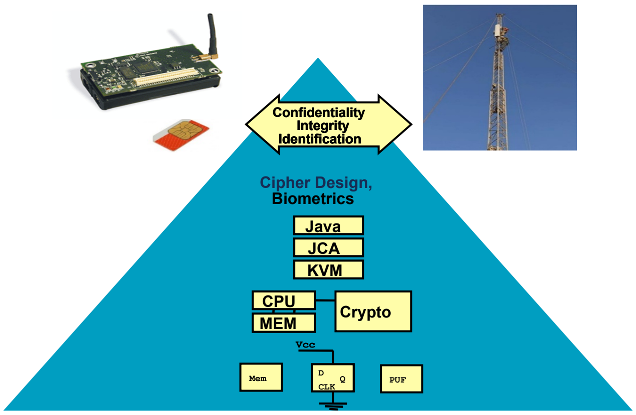
- For network system: router box
- For secure boot: the TPM or SE
- For OS designer: the architecture/micro-architecture of a processor
- For cryptographer: the VHDL or Verilog code on FPGA
- For IOT devices: attack resistance (side-channel, fault, manipulation, etc.)
- For digital designer: the standard cells or the technology

8

KU LEUVEN

HOW: DESIGN METHOD

DECOMPOSE IN COMPONENTS



- Application: secure communication
- Cryptography: public key, secret key, post-quantum,
- Architecture: Hardware/Software platform, Sancus
- Micro-architecture: crypto co-processors, instruction set extension,
- Logic circuits and (secure) memory
- TRNGs and PUFs
- Technology

[DATE2007]

"A root of trust is a component at a ***lower*** *abstraction layer, upon which the system relies for its security."*

9

KU LEUVEN

President's Council of Advisors on Science and Technology

(e) Semiconductors and System Security

Criminal and state-sponsored cyber-attacks pose increasing threats to the United States. To enable the implementation of secure systems, every aspect of the system must be considered including sensors, data converters, computing, memory, storage, and communications, while providing

REPORT TO THE PRESIDENT

Revitalizing the U.S.

We envision a research agenda in this area that should include, but is not limited to, the following: (1) design for fully secure end-to-end hardware and software solutions that are secure against various forms of attacks on operation, data, and communications; (2) security in the chip design tool chain that would enable end-to-end security solutions to be verified by design; (3) secure hardware supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

[SOURCE: Joachim Kunkel, Synopsys, Montreux 2023]

SYNOPSYS

Montreux 2023 - The Changing Semiconductor Industry Landscape 230420 - 2

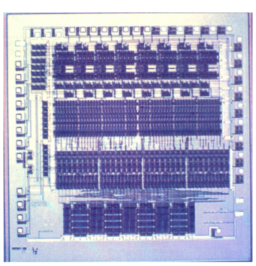
We envision a research agenda in this area that should include, but is not limited to, the following: (1) design for fully secure end-to-end hardware and software solutions that are secure against various forms of attacks on operation, data, and communications; (2) security in the chip design tool chain that would enable end-to-end security solutions to be verified by design; (3) secure hardware supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

11

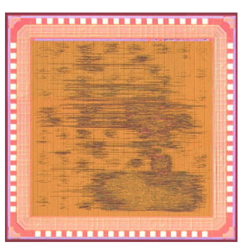
KU LEUVEN

FEASIBILITY:

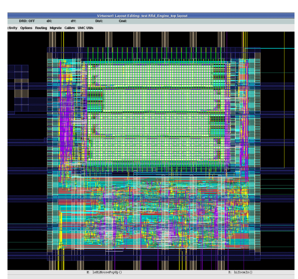
- Feasibility: what is feasible, throughput, latency, power (cooling), energy (battery lifetime) etc.



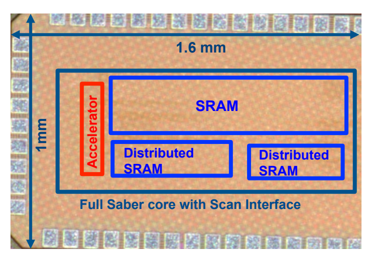
DES



Rijndael / AES



Elliptic Curve Cryptography



Post-Quantum Saber

- Next: light weight crypto, post-quantum crypto, COED

12

KU LEUVEN

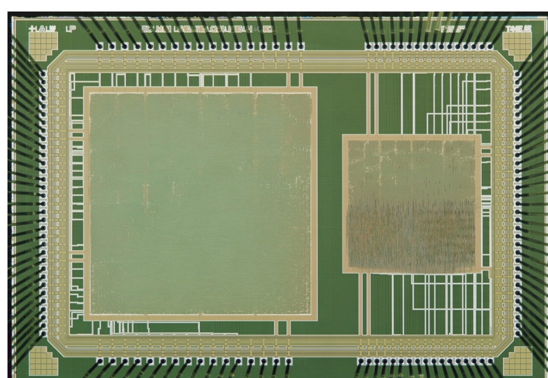
We envision a research agenda in this area that should include, but is not limited to, the following: (1) design for fully secure end-to-end hardware and software solutions that are secure against various forms of attacks on operation, data, and communications; (2) security in the chip design tool chain that would enable end-to-end security solutions to be verified by design; (3) secure hardware supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

13

KU LEUVEN

Side-channel and fault attacks

- Many types of side-channel analysis
  - Power, Electro Magnetic (EM), Time,
  - Micro-architectural side-channel: cache, transient execution attacks
- Many types of fault or active attacks:
  - EM, laser, clock, voltage glitch, etc.
- Local or remote
- Combined attacks



AES with and without countermeasure;
WDDL countermeasure integrated into
standard cell design methodology

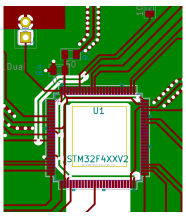
14

Measurement methods

Contact power measurements:
-shunt resistors
-current probes

Cost: 150- 5000€

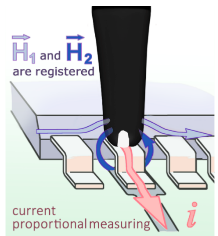
Freq: kHz – MHz range



Contactless power measurements:
-EM probes

Cost: 2000 - 25000€

Freq: kHz – GHz range

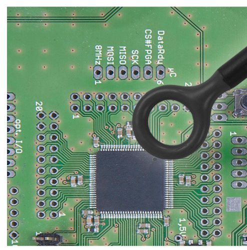


[picture credit: Langer]

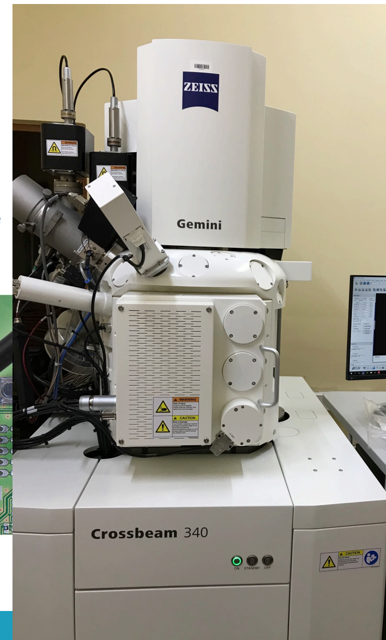
EM measurements:
-EM probes

Cost: 2000 - 25000€

Freq: kHz – GHz range



[picture credit: Langer]



Research challenges for cryptography

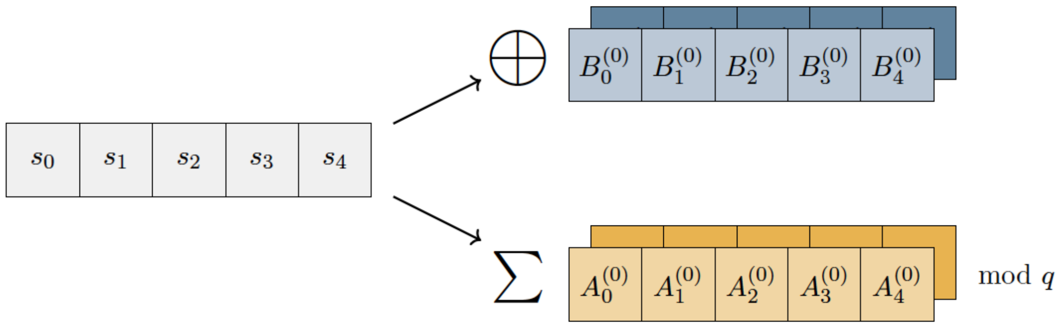
- Goal: introduce new research topics, improve existing ones

- **Challenge 1: masking is hard in practice**
- Challenge 2: PQ masking is expensive
- Challenge 3: Possibilities of PUFs
- Challenge 4: Random number generation
- Challenge 5: COED - Computing on encrypted datea

16

KU LEUVEN

Countermeasure: masking



Masking is popular in crypto community = mathematical technique

17

Countermeasure: masking

- Types of masking
  - Boolean
  - Arithmetic
  - Inner product
  - Threshold
  - …

All start from similar leakage MODEL: Shares leak independently

All require randomness

- Two experiments:
  - Symmetric key: AES masking on micro controllers
  - Public key: Post-quantum masking of lattice based encryption

18

Masking in practice is HARD

- Experiment: **first** order SW masked AES evaluated for:
  - Side-channel leakage
  - Timing
  - Randomness requirements

Paper title	Published venue	masking method
Provably Secure Higher-Order Masking of AES	CHES 2010	boolean
Higher order masking of look-up tables	Eurocrypt 2014	boolean
All the AES You Need on Cortex-M3 and M4	SAC 2016	boolean
Consolidating Inner Product Masking	Asiacrypt 2017	inner product
First-Order Masking with Only Two Random Bits	CCS-TIS 2019	boolean
Side-channel Masking with Pseudo-Random Generator	Eurocrypt 2020	boolean
Detecting faults in inner product masking scheme	JCEN 2020	inner product
Fixslicing AES-like Ciphers	TCHES 2021	boolean

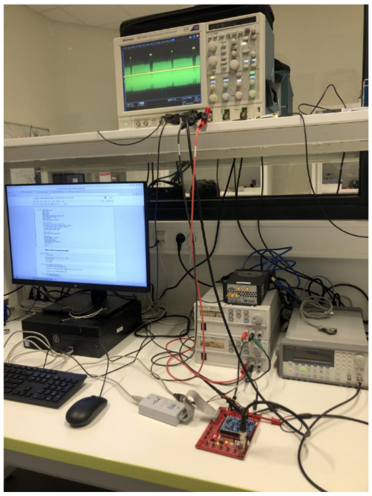
[A. Becker, L. Wouters, Cosade 2022]

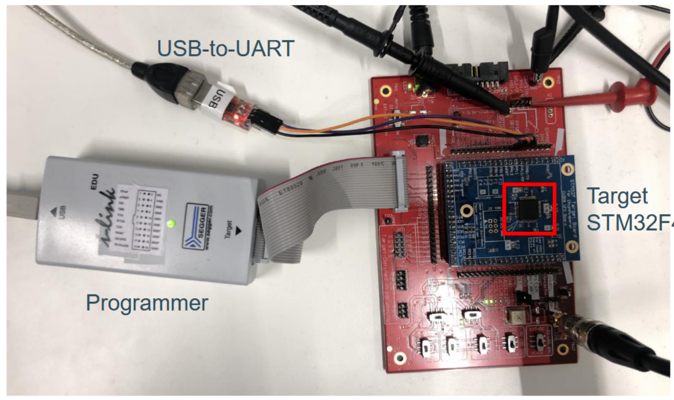
Results [Cosade 2022]

- Key recovery with first order attack
- Incorrect TRNG instantiations
- Benchmarking issues
- Software bugs

Paper title	Published venue	masking method
Provably Secure Higher-Order Masking of AES	CHES 2010	boolean
Higher order masking of look-up tables	Eurocrypt 2014	boolean
All the AES You Need on Cortex-M3 and M4	SAC 2016	boolean
Consolidating Inner Product Masking	Asiacrypt 2017	inner product
First-Order Masking with Only Two Random Bits	CCS-TIS 2019	boolean
Side-channel Masking with Pseudo-Random Generator	Eurocrypt 2020	boolean
Detecting faults in inner product masking scheme	JCEN 2020	inner product
Fixslicing AES-like Ciphers	TCHES 2021	boolean

Set-up in the lab



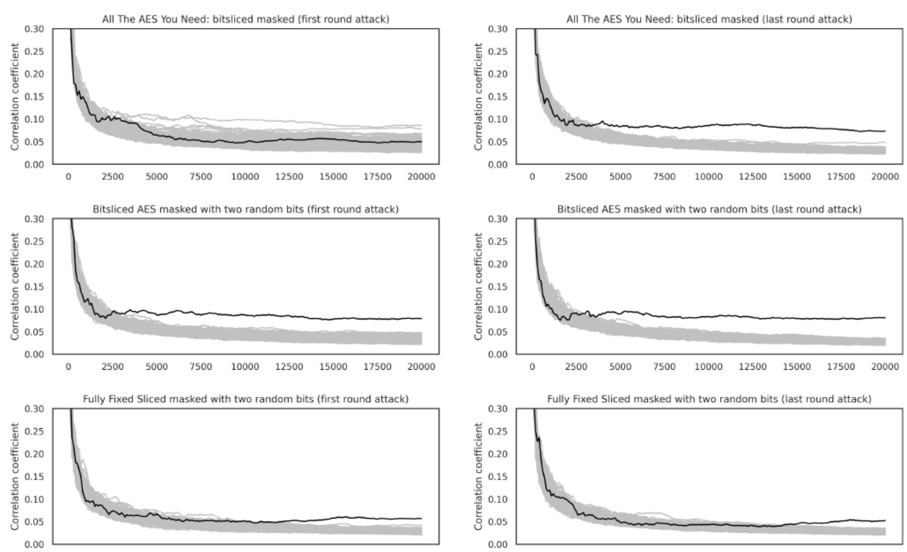


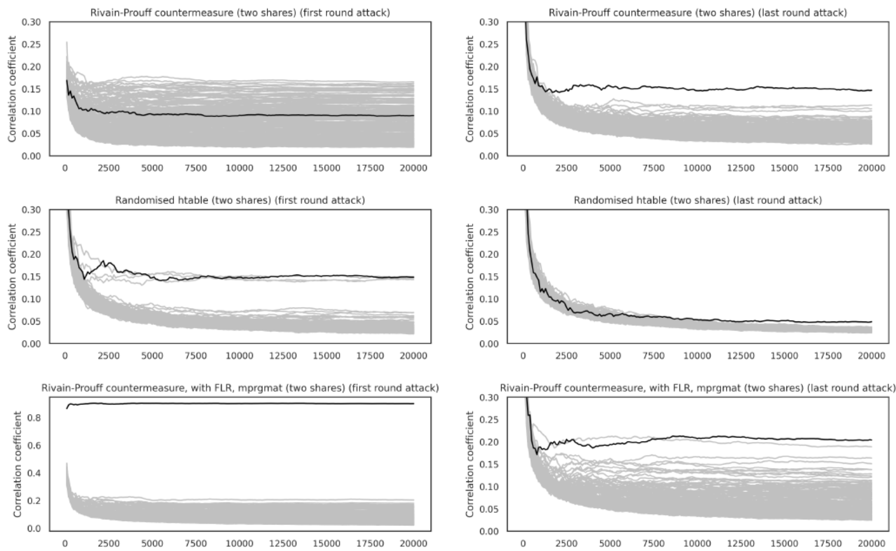
21

CPA results

AES Key recovery:
- Byte by byte
- correct byte stands out to 255 other options

X-axis: number of samples
Y-axis: correlation coef





For more details see COSADE 2022

Cause: violation of assumptions

- Assumption: shares leak independently
- Leakage caused by the microcontroller breaks this assumption
  - Assume share A is in r0
  - Move share B into r0 (and overwrite share A)
  - Information on A ⊕ B is leaked!

EDA message:
TOOLS could help here!

- Complex processors: transient execution
- Compiler optimizations
- Coupling through power and ground network
- Below 60nm CMOS ‘static’ leakage

Research challenges for cryptography

- Goal: introduce new research topics, improve existing ones

- Challenge 1: masking is hard in practice
- **Challenge 2: Post quantum masking is expensive**
- Challenge 3: Possibilities of PUFs
- Challenge 4: Random number generation
- Challenge 5: NEW – Fully Homomorphic Encryption

25

KU LEUVEN

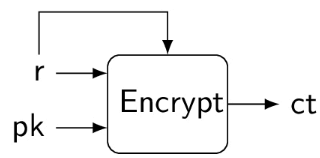
We envision a research agenda in this area that should include, but is not limited to, the following: (1) design for fully secure end-to-end hardware and software solutions that are secure against various forms of attacks on operation, data, and communications; (2) security in the chip design tool chain that would enable end-to-end security solutions to be verified by design; (3) secure hardware supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

26

KU LEUVEN

Lattice Based Post-quantum crypto (NIST)

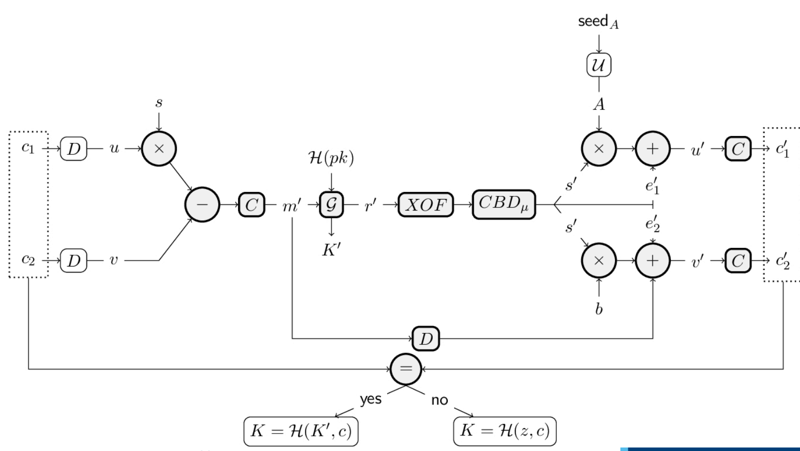
- KEM = key generation, encapsulation, decapsulation
- CCA secure: Fujisaki – Okamoto transformation
- Similar for
  - Kyber
  - Saber



KU LEUVEN

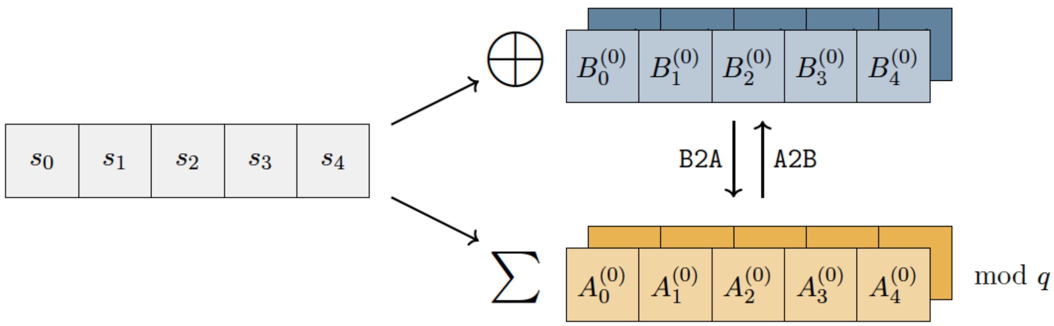
Cost of decapsulation

- Expensive parts: multiplication, hash, sampling
- Saber vs Kyber
  - Very similar
  - Power of two $q = 2^{13}$ vs $q = 3329$
  - MLWR vs MLWE implicit vs explicit error addition



KU LEUVEN

Arithmetic and Boolean masking



Conversion is: Arithmetic to Boolean (A2B) or Boolean to Arithmetic (B2A)

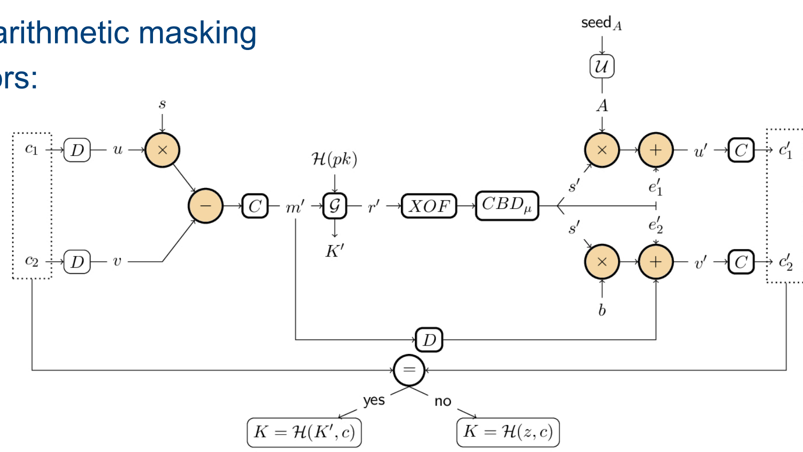
29

KU LEUVEN

Polynomial arithmetic

- Easy to protect with arithmetic masking
- Small overhead factors:
  - 1.7 to 2.0 (n=2)
  - 2.96 (n=3)

n = sharing factor



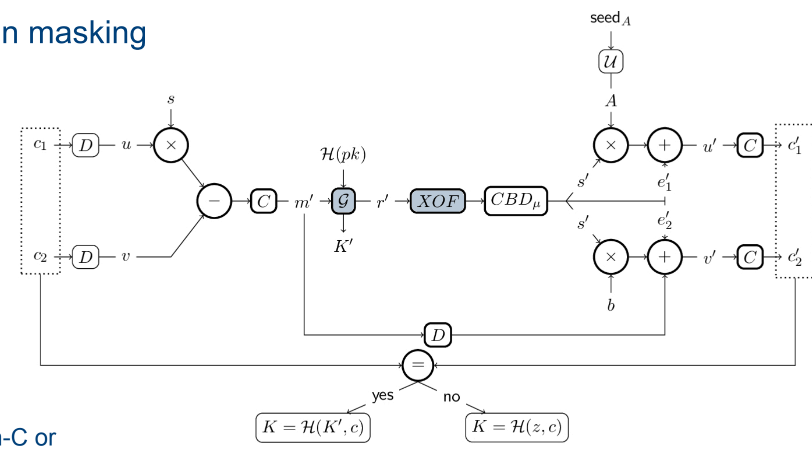
[BGR+21, BDK+21]

30

KU LEUVEN

SHA-3

- Protected with Boolean masking
- Overhead factors
- 5.9 to 9.26 (n = 2)
- 73.1 (n=3)



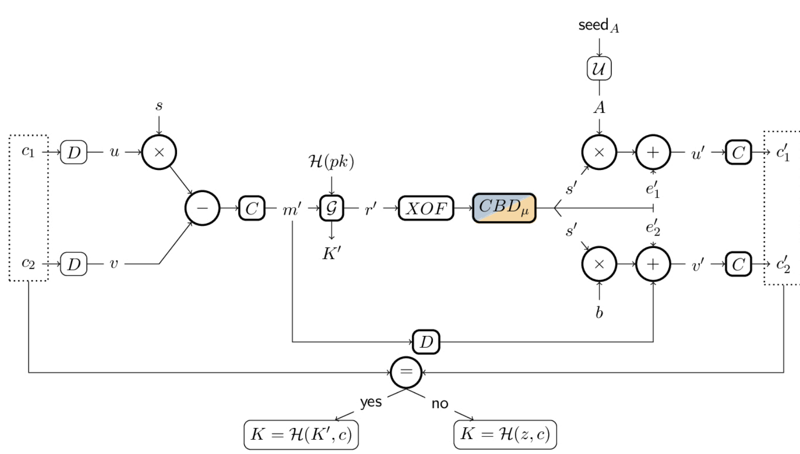
Depends if you compare to plain-C or optimized assembly

[Boolean masking: BDPVA10,BBD+16]

KU LEUVEN

Centered Binomial sampling

- Mix of A2B and B2A
- Expensive!
- Etc.

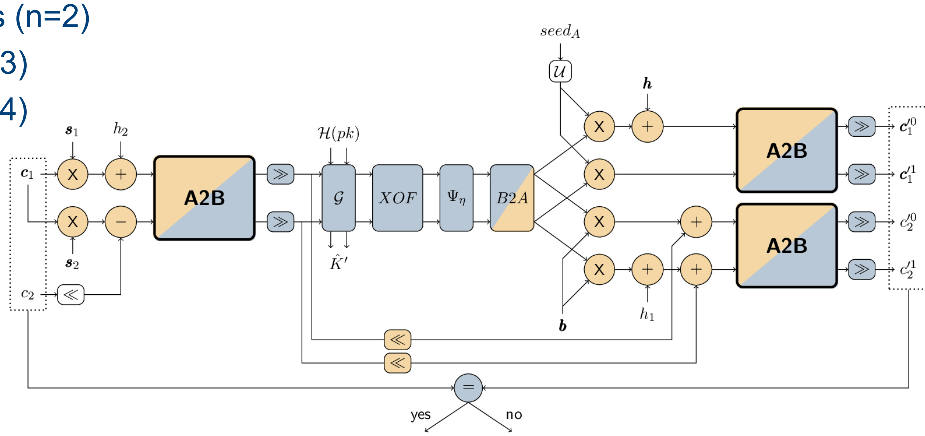


KU LEUVEN

One A2B conversion cost (Saber)

Requires bit-slicing
- 55-61 K cycles (n=2)
- 172-206 K (n=3)
- 302-365 K (n=4)

\+ randomness



[ABV22, BC22]

Masking is expensive

CPU cycles x1000 / Scheme	Unmasked	1st order n=2	2nd order n=3	3rd order n=4
Saber	773	3,011 (1x)	5,534 (1x)	8,591 (1x)
Kyber [2]	804	7,716 (2.56x)	11,880 (2.14x)	16,715 (1.94x)
COST	**1x**	**3.9x – 9.6x**	**7.2x – 14.8x**	**11.1x – 20.8x**
Random bytes		**12 KB**	**42 KB**	**90 KB**

- Unmasked Kyber/Saber similar cost
- Masked Kyber more expensive vs Saber
  - Power of two
  - Rounding vs error sampling
- Masking is expensive AND requires randomness

Platform: ARM Cortex M4
Framework: PQM4
Compiled: arm-none-eabi-gcc
Version: 9.2.1

Research challenges for cryptography

- Goal: introduce new research topics, improve existing ones

- Challenge 1: masking is hard in practice
- Challenge 2: PQ masking is expensive
- Challenge 3: Possibilities of PUFs
- Challenge 4: Random number generation
- **Challenge 5: Computing on encrypted data - FHE**
  - On FPGA
  - On ASIC

35

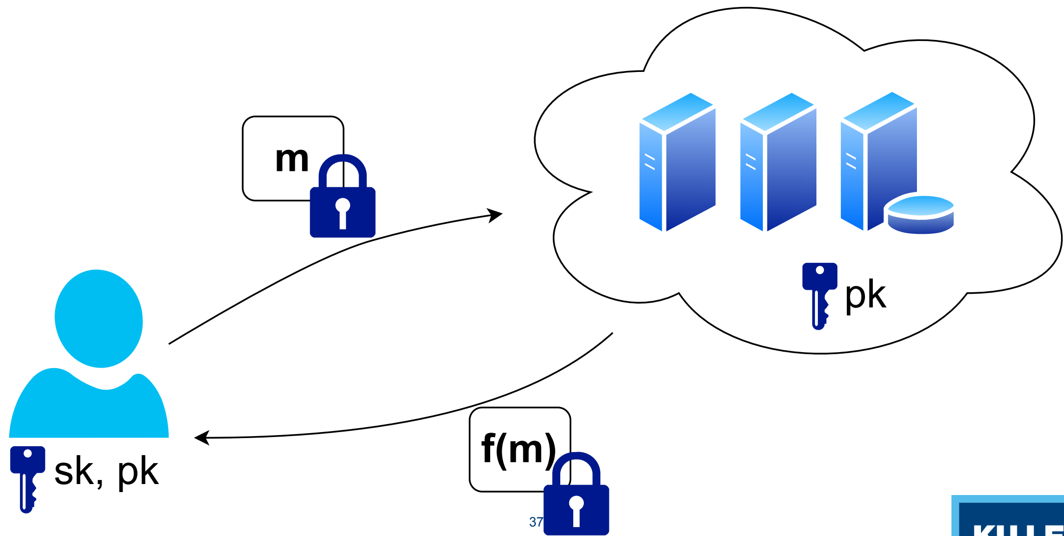
KU LEUVEN

We envision a research agenda in this area that should include, but is not limited to, the following: (1) design for fully secure end-to-end hardware and software solutions that are secure against various forms of attacks on operation, data, and communications; (2) security in the chip design tool chain that would enable end-to-end security solutions to be verified by design; (3) secure hardware supply chain covering chip fabrication, packaging, and system integration; (4) implementation of post-quantum cryptography; (5) implementation of low-power cryptography for secure communications and transactions; and (6) other privacy preserving hardware implementations for processing encrypted data.

36

KU LEUVEN

Fully Homomorphic Encryption



37

KU LEUVEN

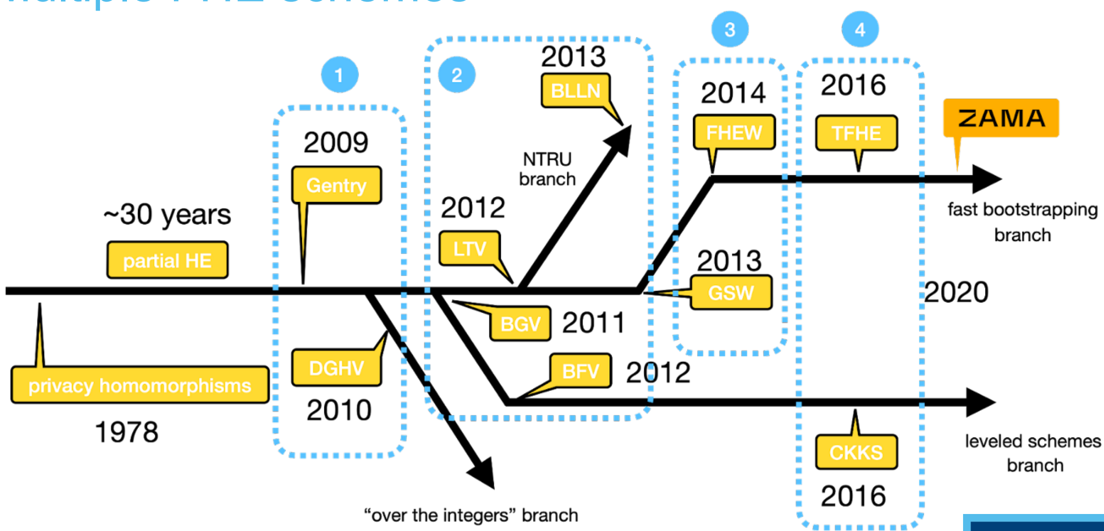
Multiple schemes

- Partially homomorphic: Paillier system
- Somewhat homomorphic:
  - Limited number of multiplications
  - Fan-Vercauteren:
- Fully Homomorphic Encryption
  - Unlimited number of multiplications
  - Requires 'bootstrapping'
- Multiple schemes:
  - BFV: Brakerski – Fan – Vercauteren
  - BGV: Brakerski – Gentry – Vaikuntanathan
  - …

38

KU LEUVEN

Multiple FHE schemes



[copied from ZAMA website]

39

KU LEUVEN

Challenge large numbers:

- Experiment 1 [CHES2015] : YASHE (now no longer used, reduced security)
  - Ciphertext size 5MB to 20MB (Polynomial size is 32768 (2^15) to 65536 (2^16), modulus 1200 to 2500 bits), could evaluate depth of Simon block cipher
- Experiment 2 [TC2018]: HEPCLOUD, FV
  - Ciphertext pair 9.2MB with parameters Polynomial size is 32768 (2^15), modulus 1128 bits, depth 36, 85 bits security level.
  - Bottleneck: I/O between FPGA and external memory
- Experiment 3 [TC2020]: HEAWS, FV
  - Cipher text pair 180KB, with parameters Polynomials size is 4096, modulus min 372 (Q), 180 (q), depth 4, more than 80 bits security.
  - Useful for small neural network applications
  - Fits on one FPGA

40

KU LEUVEN

DARPA DPRIVE program: in progress

- Dedicated ASIC acceleration of BGV
  - 150mm^2 chip in 12nm GF
  - Within 10× of plaintext computation
  - 10,000× faster than software reference
  - Parameter set for 128-bit security
    - Support *bootstrapping*
- Four teams of researchers
  - **Galois**, Duality, SRI, and Intel
- Several phases
  - **Phase 1**: design, implementation and verification of system architecture and IP blocks

Now: phase 2 running, with three teams: Galois, Duality and Intel

41

KU LEUVEN

BGV parameters in DPRIVE

Parameter	Range	Example
Security parameter	N/A	128 bits
Ring dimension N	$512 - 65536$	$\mathbf{65536}$
Plaintext modulus p^r	$\geqslant 2$	127^3
Ciphertext packing ℓ	$1 - 65536$	64 slots
Max $\log_2(QP)$ for key switching	$20 - 1782$	**1782 bits**
Max $\log_2(Q)$ for ciphertext	$20 - 1263$	1263 bits
Max multiplicative depth L	N/A	31

Ciphertext: 21 MB, Key-switch key: 84 MB

42

KU LEUVEN

Hardware acceleration options

43

KU LEUVEN

Challenges

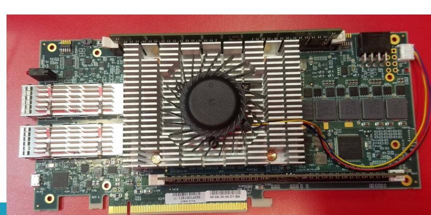
- Computational complexity
  - NTT/FFT acceleration
- **Memory**
  - SIZE
  - BANDWIDTH

ASIC (phase 1)

- 150 mm^2 in 12nm
- Global Foundaries
- Memory hierarchy
- 57 – 115 Watt

Cloud FPGA

- Alveo U280 (in 5nm or 7 nm)
- Included into Amazon AWS F1
- Memory hierarchy
- 225 Watt! (cooling)



44

KU LEUVEN

Three experiments – two domain specific processors

FPGA - HEAWS

- BFV – leveled HE
- 80 bit security
- Shallow depth

ASIC – DPRIVE – BASALISC

- BGV – includes Bootstrap
- 128 bit security
- DPRIVE constraints

- NTT acceleration
- Residue Number System

- Dedicated instruction set
- No cache: compile time known

IEEE TC 2020

IACR 2022/657

45

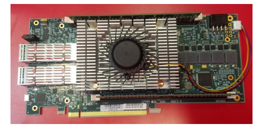
KU LEUVEN

First experiment: FPGA Acceleration of BFV on Amazon cloud

46

KU LEUVEN

FPGA Memory Resources (Alveo U280)



BRAM – 9 MB

URAM – 33 MB

LOCAL ON CHIP

HBM – 8 GB

IN PACKAGE, 3D

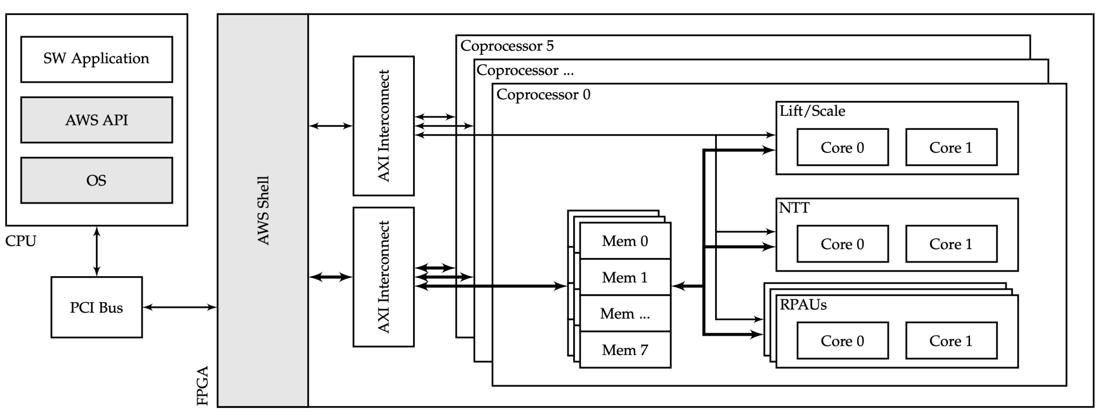
DDR – 32 GB

ON BOARD

47

KU LEUVEN

Amazon F1 Instance Implementation [TC20]



48

KU LEUVEN

Performance of Homomorphic Multiplication

- Each multiplication takes 4.34 ms.
- The overhead of a ciphertext transfer is 0.11 ms.

- A single coprocessor achieves 230 multiplications per second.
- **Six** coprocessors running in parallel achieves 613 multiplications.

Comparison

- Achieve 613 homomorphic multiplications per second
- Compared to CPU
  - 13x speedup w.r.t. a highly optimized software on Intel i5 processor, 1.8 GHz
- To GPU on Amazon cloud, **5 times more work for half price and lower power!**

Compute: Amazon EC2 Instances:

	Description	Instances	Usage		Type	Billing Option	Monthly Cost
	1 FPGA -> 2000 Mult	1	100	% Utilized/Mo	Linux on f1.2xlarge	On-Demand (No Co	$ 1207.80
	1 GPU -> 388 Mult	1	100	% Utilized/Mo	Linux on p3.2xlarge	On-Demand (No Co	$ 2239.92

14.02.2019 - AWS Simple Monthly Calculator: https://calculator.s3.amazonaws.com/index.html
17.06.2020

Second experiment: ASIC Acceleration of BGV

Darpa DPRIVE Basalisc project

51

KU LEUVEN

BASALISC Memory Hierarchy

DPRIVE
DATA PROTECTION IN VIRTUAL ENVIRONMENTS

- MAC Acc – 8 KB
- MAC Register File – 128 KB
- Cipher Text Buffer – 64 MB
- DDR – 256 GB

ON CHIP, Cipher Text Buffer
CTB fits 3 ciphertext pairs

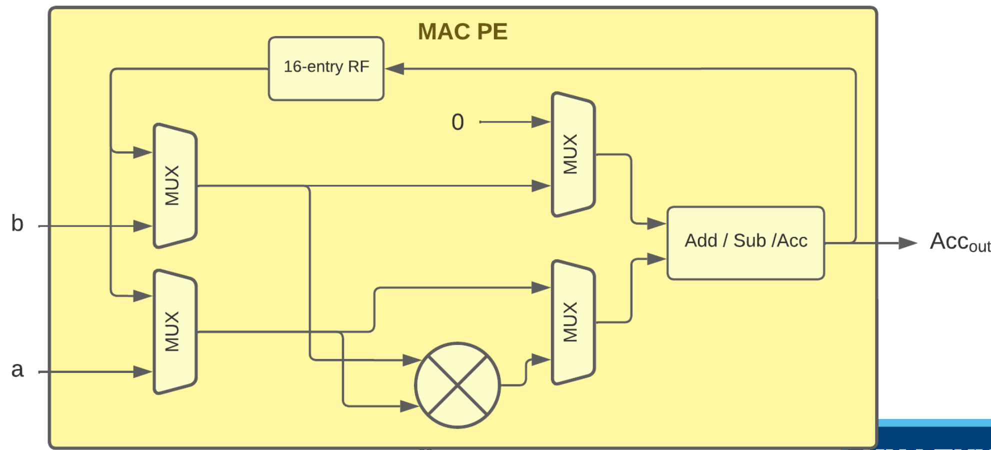
One Key switch 84MB does not fit

ON BOARD

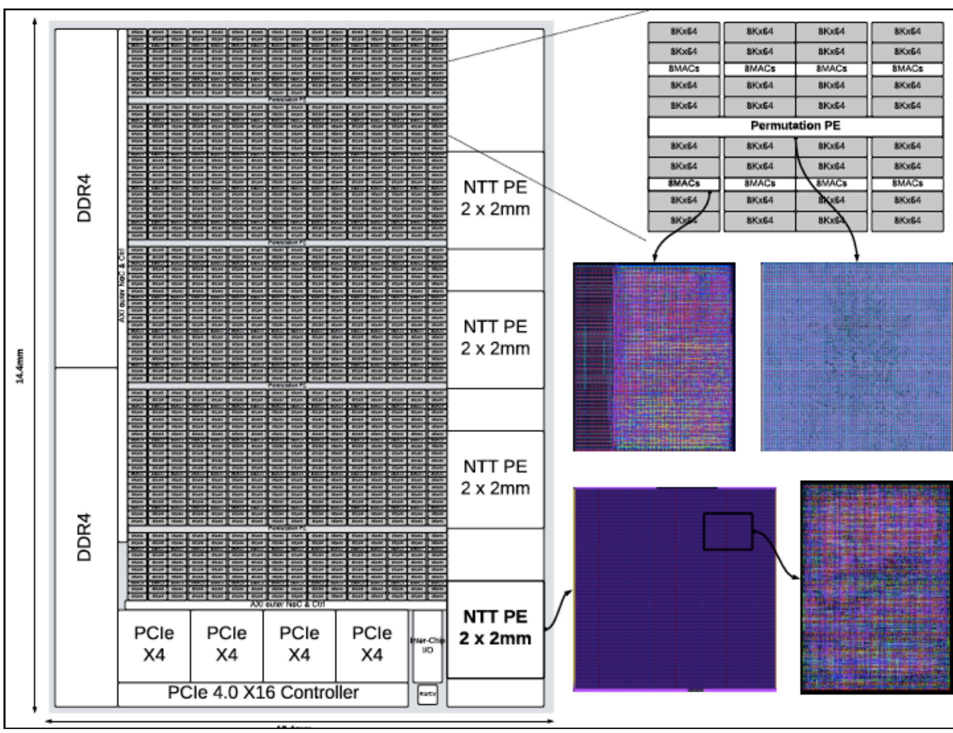
52

KU LEUVEN

2048 x 32 bits Modular Multiply – Accumulate unit PE



KU LEUVEN



MACs distributed into memory

(compute near/in memory)

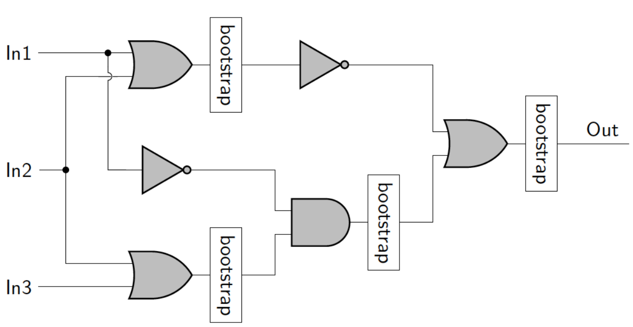
KU LEUVEN

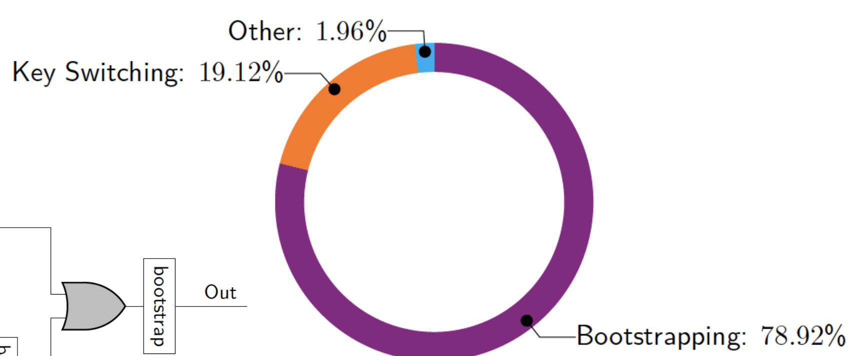
Third experiment: FPT
FPGA Fixed Point Accelerator for TFHE
Torus Fully Homomorphic Encryption

ERC Advanced Grant Belfort, FWO

KU LEUVEN

Challenge: Bootstrap acceleration

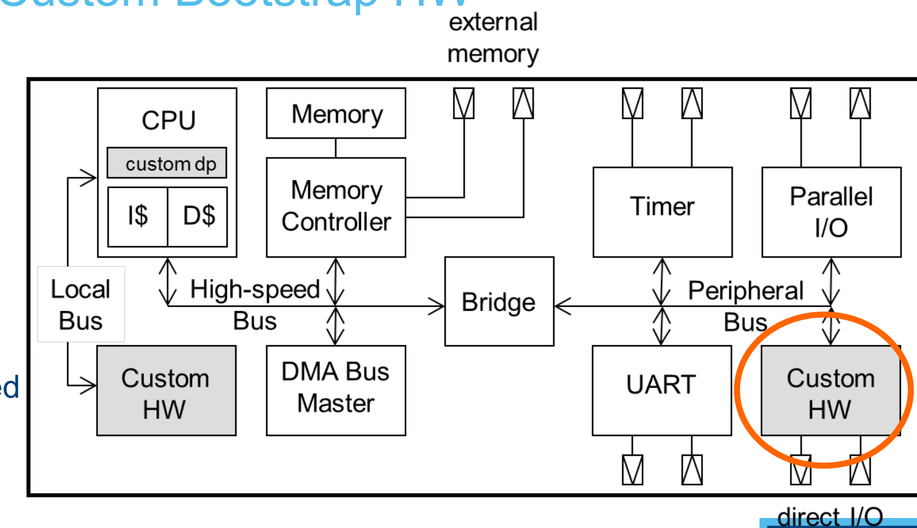




KU LEUVEN

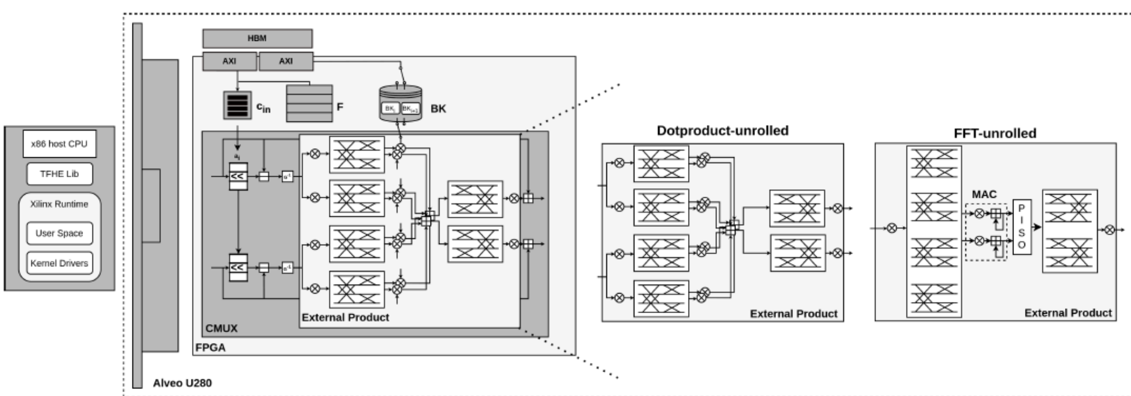
Option 3: Custom Bootstrap HW

- Peripheral
  = loosely coupled



[Picture: P. Schaumont, “A practical introduction to Hardware/Software Codesign”, 2nd ed

FPGA: Bootstrap FFT accelerator



Results

- FPGA
- ASIC
- CPU
- GPU

	LUT / FFs / DSP / BRAM	f (MHz)	l (ms)	TP (PBS/ms)
FPT	595K / 1024K / 5980 / 14.5Mb	200	0.58	**25.0**
YKP	842K / 662K / 7202 / 338Mb 442K / 342K / 6910 / 409Mb	180 180	3.76 1.88	3.5 2.7
MATCHA	36.96mm^2 16nm PTM	2000	0.2	10
CONCRETE	Intel Xeon Silver 4208	2100	32	0.03
cuFHE	NVIDIA GeForce RTX 3090	1700	9.34	9.6

FHE demonstration available from:
https://www.youtube.com/watch?v=Bbkc1lavkGo&list=PLnbmMskCVh1ei6AkXHDTAefkGZaBmtUQO&index=10
Presented at FHE 2023

59

KU LEUVEN

Conclusions – lessons learned

- Trust and trustworthy design
- Efficient AND secure cryptography
- Masking as countermeasure is hard and expensive

- Novel crypto challenges:
  - Post-quantum cryptography
  - Computing on encrypted data

60

KU LEUVEN